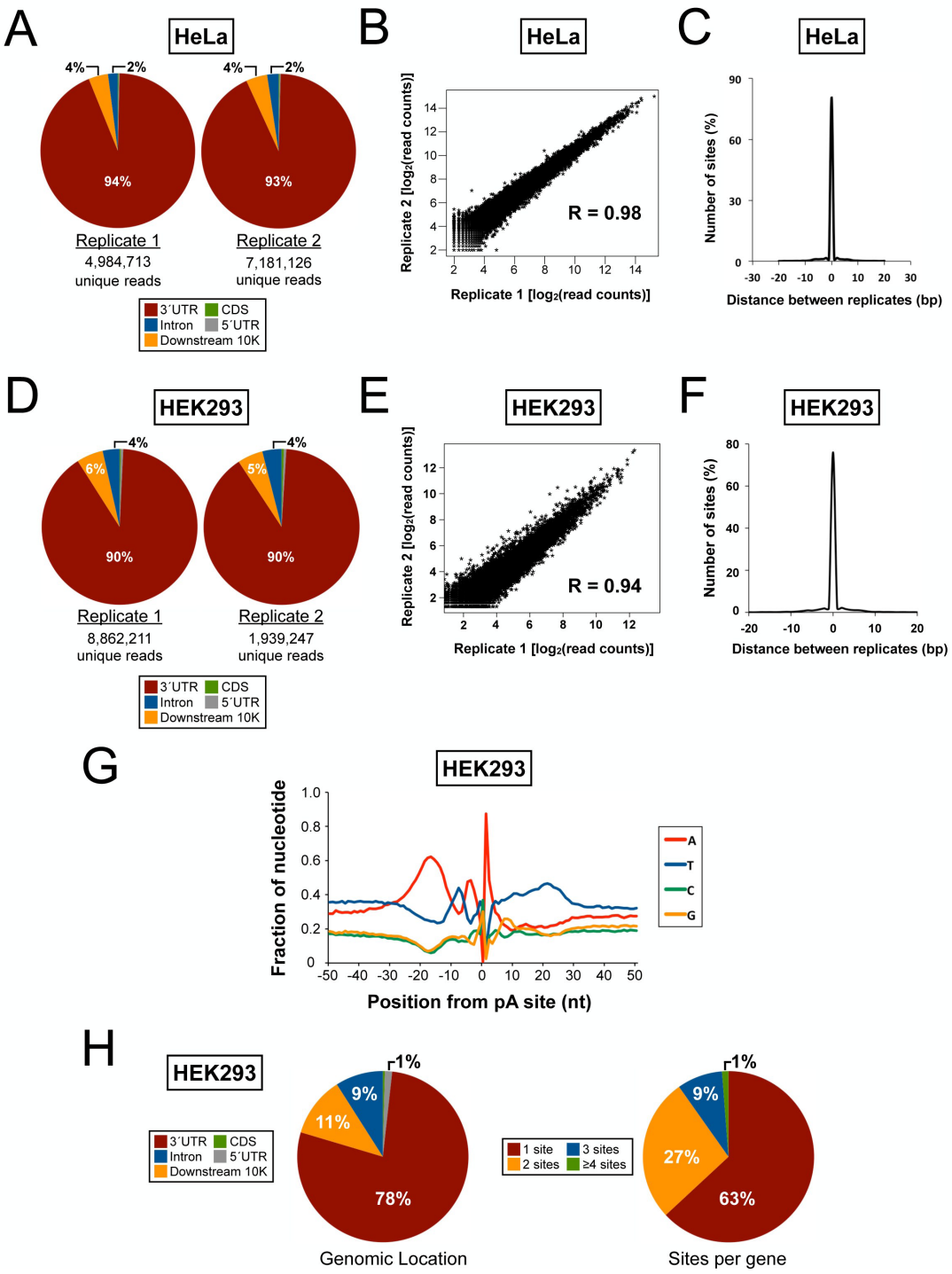

Figure S1. PAPERCLIP is highly reproducible. Related to Figure 2. (**A**) Pie-charts showing the genomic distribution of uniquely mapped reads from two PAPERCLIP replicate experiments in HeLa cells. The numbers of uniquely mapped reads are listed below the pie-charts. (**B**) A scatter plot showing the correlation of read counts at 17,652 poly(A) sites between two PAPERCLIP replicates. Each dot represents a poly(A) site. R, Pearson's correlation coefficient. (**C**) A diagram showing the distance between assigned positions for 17,652 poly(A) sites between two PAPERCLIP replicates. 98% of the sites are within ±10 bp between replicates. (**D**) Pie-charts showing the genomic distribution of uniquely mapped reads from two PAPERCLIP replicate experiments in HEK293 cells. The numbers of uniquely mapped reads are listed below the pie-charts. (**E**) A scatter plot showing the correlation of normalized read counts at 16,414 poly(A) sites between two PAPERCLIP replicates. Each dot represents a poly(A) site. R, Pearson's correlation coefficient. (**F**) A diagram showing the distance between assigned positions for 16,414 poly(A) sites between two PAPERCLIP replicates. 97% of the sites are within ±10 bp between replicates. (**G**) A diagram showing the genomic nucleotide sequence surrounding the 16,414 poly(A) sites identified by PAPERCLIP in HEK293 cells. (**H**) Pie-charts showing: the genomic location (left) and the per gene distribution (right) of poly(A) sites identified by PAPERCLIP in HEK293 cells.

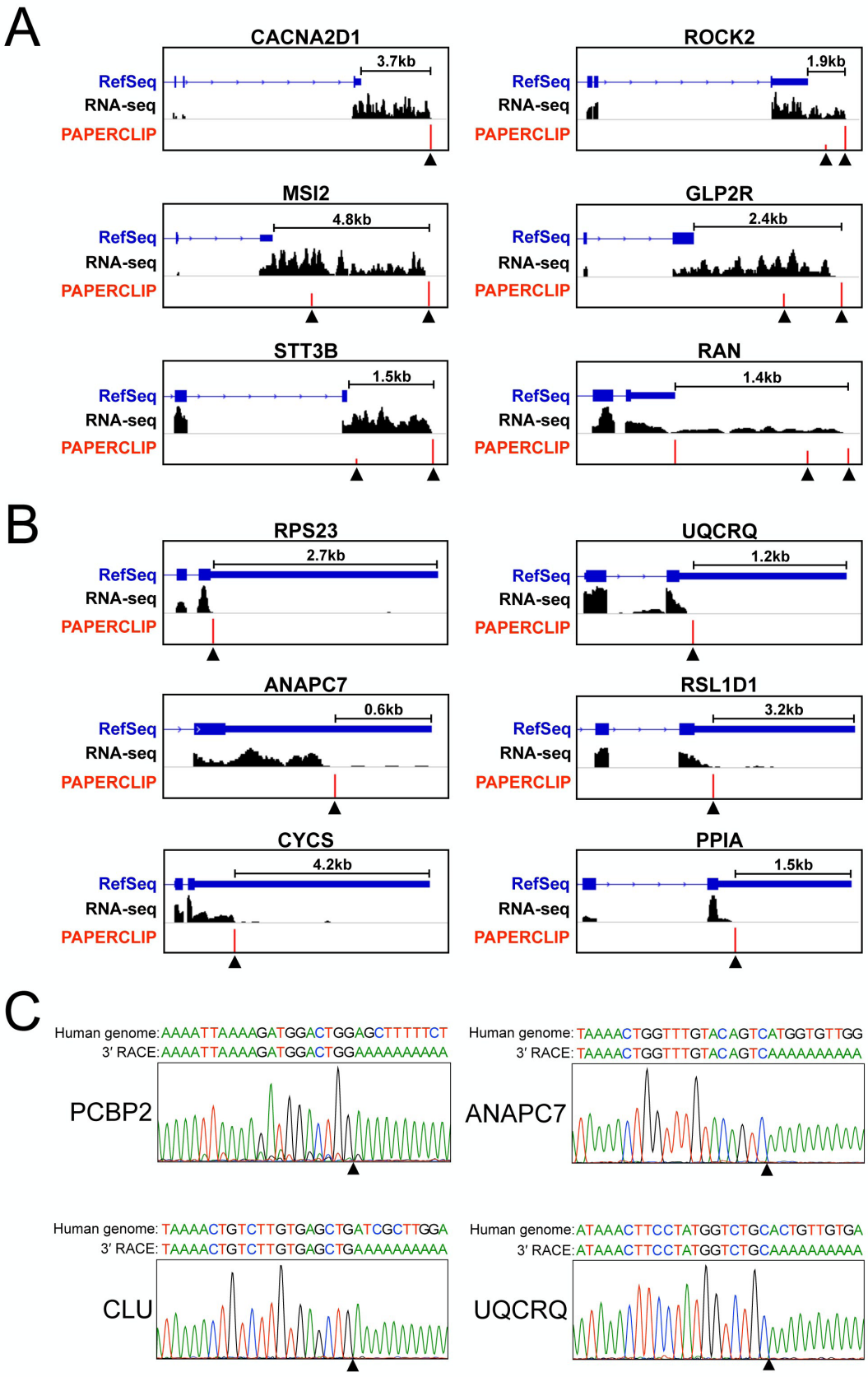

Figure S2. RNA-seq and 3′ RACE support for PAPERCLIP annotation in HeLa cells. Related to Figure 2. (**A**) and (**B**) Diagrams showing RefSeq annotation, RNA-seq and PAPERCLIP results. The peak heights from PAPERCLIP are scaled for each gene. The bar above RefSeq annotation indicates the length of 3′ UTR extension (**A**) or shortening (**B**). Arrowheads indicate poly(A) sites not present in the RefSeq annotation. (**C**) Diagrams showing sequencing traces from 3′ RACE experiments and human genome sequence for comparison. Arrowheads indicate poly(A) sites identified by PAPERCLIP.

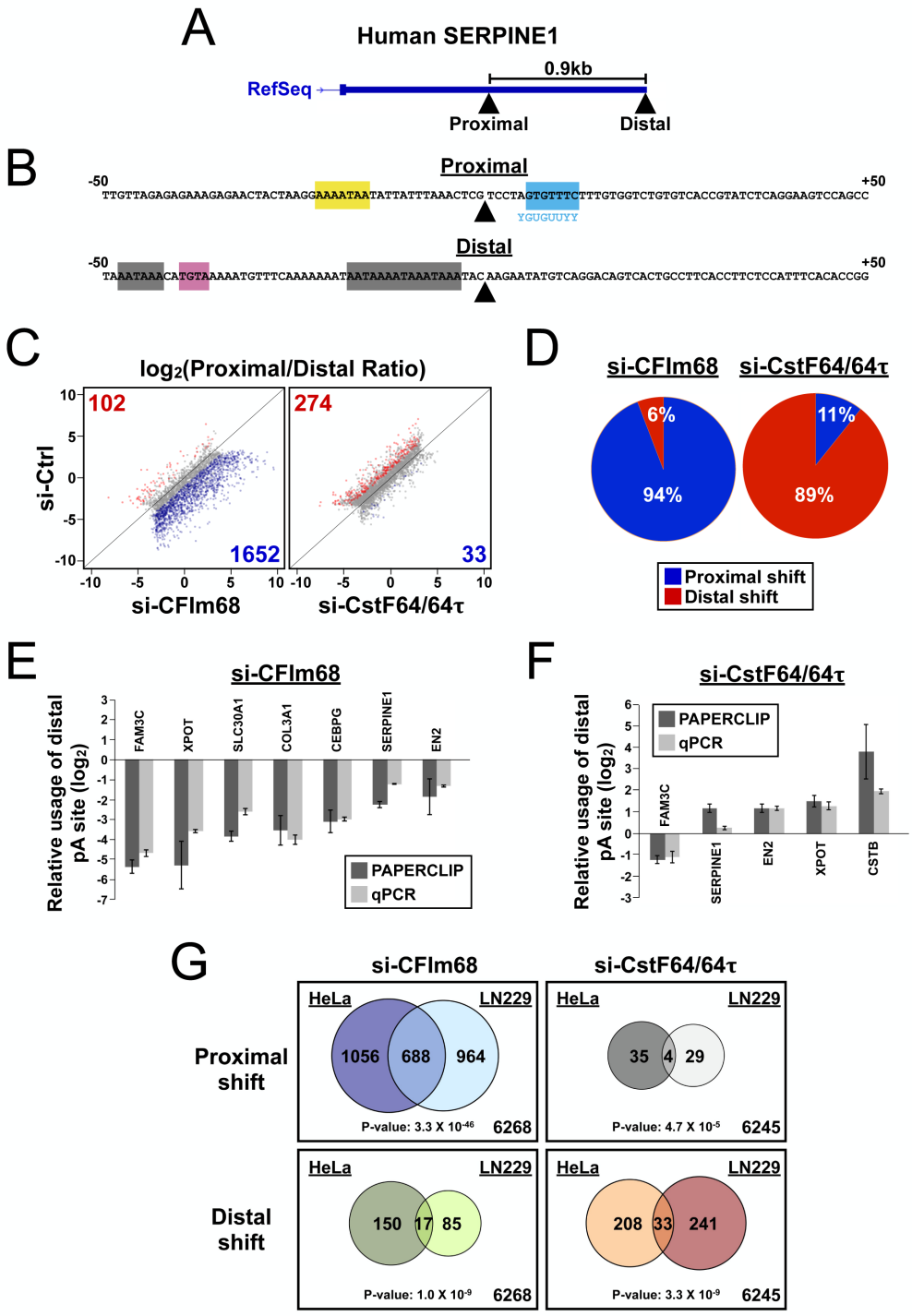

Figure S3. PAPERCLIP identifies APA shifts in LN229 cells. Related to Figure 3. (**A**) A diagram showing the RefSeq annotation for the last exon of SERPINE1 and the locations of the proximal and distal poly(A) sites. (**B**) Diagrams showing both the upstream 50bp and downstream 50bp genomic nucleotide sequence of SERPINE1 proximal and distal poly(A) sites. Black triangles denotes the poly(A) sites. Known motifs are color-shaded as in Figure 3G. A consensus of mammalian DSE, YGUGUUYY, is depicted below the DSE of proximal poly(A) site for comparison. (**C**) Scatter plots comparing $\log_2$(proximal/distal ratio) by PAPERCLIP for 2-peak genes between control siRNAs (si-Ctrl) and treatment siRNAs (si-CFIm68 and si-CstF64/64tau). Each dot represents a gene. Genes with FDR<0.05 and at least twofold change of P/D ratio are considered significantly shifted and are colored. Red, $\log_2$[(treatment siRNA P/D ratio)/(control siRNA P/D ratio)]≥1. Blue, $\log_2$[(treatment siRNA P/D ratio)/(control siRNA P/D ratio)]≤-1. Total numbers of significantly shifted genes for both directions are listed at the corners of plots. (**D**) Pie-charts summarizing the direction of APA shift in significantly shifted genes from (**C**). (**E**) and (**F**) Diagrams comparing PAPERCLIP results from (**C**) and validating qRT-PCR experiments for individual genes. Error bars represent standard errors. (**G**) Venn diagrams summarizing the overlap between HeLa and LN229 siRNA experiments. The overlap between HeLa and LN229 in all four groups is statistically significant by hypergeometric test.

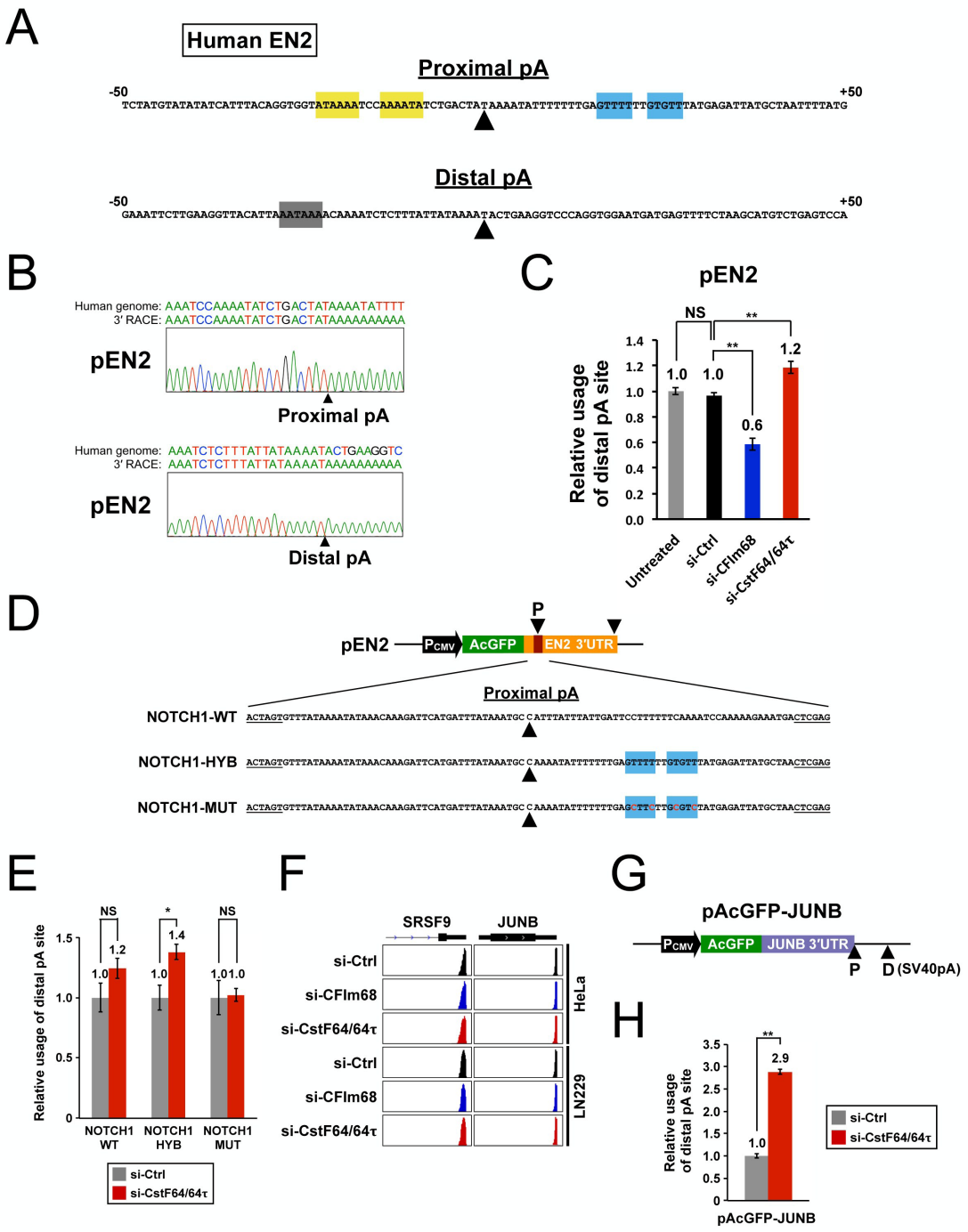

Figure S4. pEN2 provides an experimental tool to study non-canonical poly(A) site usage. Related to Figure 4. (**A**) Diagrams showing both the upstream 50bp and downstream 50bp genomic nucleotide sequence of EN2 proximal and distal poly(A) sites. Black triangles denotes the poly(A) sites. Sequence motifs are color-shaded as in Figure 3G and S5. Blue, GUKKU motif. Yellow, adenine-rich sequence. Gray, canonical poly(A) signal. (**B**) Diagrams showing sequencing traces from 3′ RACE experiments and human genome sequence for comparison. Arrowheads indicate poly(A) sites identified by PAPERCLIP. (**C**) Bar graphs showing qRT-PCR results from two independent siRNA experiments. Error bars represent standard errors. **: p<0.01. NS: not significant. (**D**) Diagrams showing the nucleotide sequence of inserted poly(A) sites from NOTCH1 proximal poly(A) site and its derivatives flanked by restriction sites for cloning (underline). Blue, GUKKU motif. (**E**) Bar graphs showing qRT-PCR results from two independent siRNA experiments. Error bars represent standard errors. *: p<0.05. NS: not significant. (**F**) Diagrams showing the last exon of SRSF9 and the entire JUNB gene, and PAPERCLIP read clusters in three experimental conditions in HeLa and LN229 cells. (**G**) A diagram showing the pAcGFP-JUNB construct. Black triangles denotes the poly(A) sites. The original SV40 poly(A) site located downstream of the multiple cloning site serves as the distal poly(A) site. (**H**) Bar graphs showing qRT-PCR results from two independent siRNA experiments. Error bars represent standard errors. **: p<0.01.

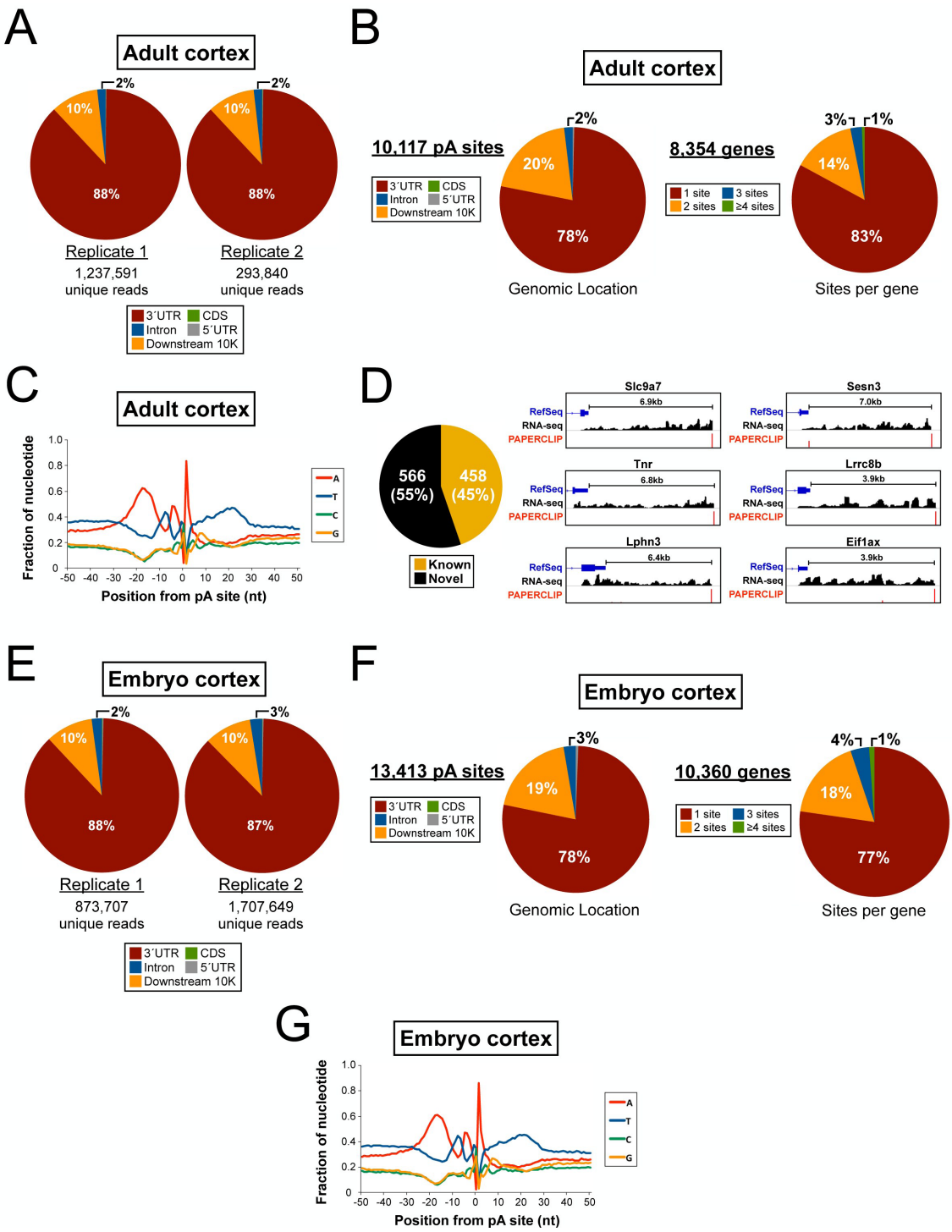

Figure S5. PAPERCLIP identifies novel 3′ UTR extensions in adult mouse cortex. Related to Figure 5. (**A**) Pie-charts showing: the genomic location (left) and the per gene distribution (right) of poly(A) sites identified by PAPERCLIP in adult mouse cortex. (**B**) A diagram showing the genomic nucleotide sequence surrounding the 10,117 poly(A) sites identified by PAPERCLIP in adult mouse cortex. (**C**) Pie-charts showing: the genomic location (left) and the per gene distribution (right) of poly(A) sites identified by PAPERCLIP in adult mouse cortex. (**D**) (left) A pie-chart showing the numbers and percentage of known and novel 3′ UTR extensions identified by PAPERCLIP. (right) Diagrams showing RefSeq annotation, RNA-seq and PAPERCLIP results. The bar above RefSeq annotation indicates the length of 3′ UTR extension. (**E**) Pie-charts showing the genomic distribution of uniquely mapped reads from two PAPERCLIP replicate experiments in embryonic mouse cortex. The numbers of uniquely mapped reads are listed below the pie-charts. (**F**) Pie-charts showing: the genomic location (left) and the per gene distribution (right) of poly(A) sites identified by PAPERCLIP in embryonic mouse cortex. (**G**) A diagram showing the genomic nucleotide sequence surrounding the 13,413 poly(A) sites identified by PAPERCLIP in embryonic mouse cortex.

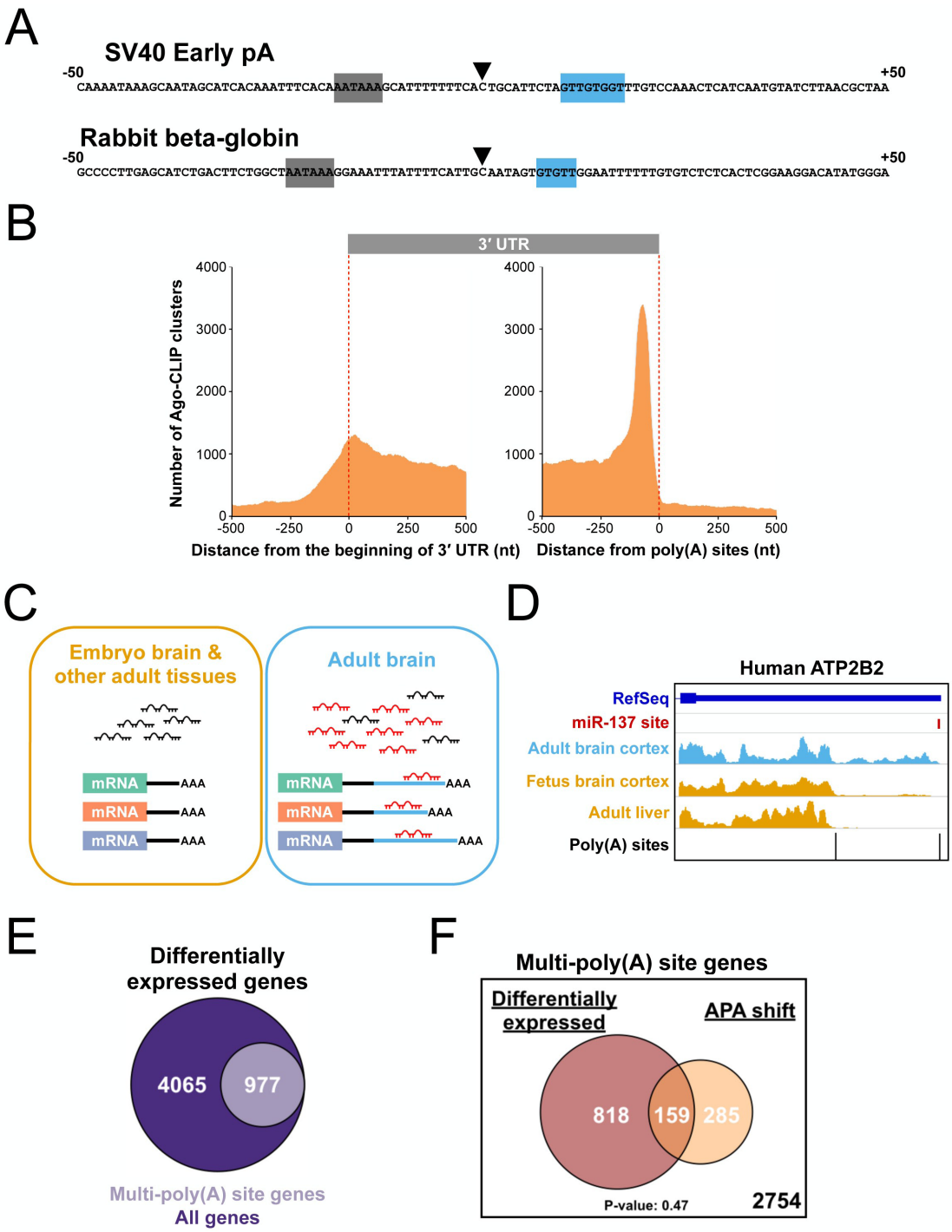

Figure S6. PAPERCLIP data provides insights into gene regulation in mouse brain development. Related to Figure 6. (**A**) Diagrams showing both the upstream 50bp and downstream 50bp genomic nucleotide sequence of the SV40 early poly(A) site and rabbit beta-globin poly(A) site. Black triangles denotes the poly(A) sites. Blue, GUKKU motif. Gray, the canonical poly(A) signal. (**B**) Diagrams showing the collective spatial distribution of Ago-CLIP clusters relative to the beginning of 3′ UTR (left) or PAPERCLIP identified poly(A) sites (right). (**C**) Schematics showing both the expression of neuron-enriched miRNAs (depicted in red) and neural 3′ UTR extensions (light blue) in adult brain (right) constitute a gene regulatory circuit that is not present in embryo brain or other adult tissues (left) in *Drosophila* and in mammals. Widely expressed miRNAs and the ubiquitous 3′ UTR sequences are depicted in black. AAA represents the poly(A) tail. (**D**) Diagrams showing the RefSeq annotation, the miR-137 site, adult and fetus brain RNA-seq results from Jaffe et al., adult liver RNA-seq result from the ENCODE project (ENCFF014GVA) and poly(A) sites annotated by PAPERCLIP for human *ATP2B2*. Except for adult liver RNA-seq, all other tracks are re-plotted from Fig. 6D for comparison. (**E**) A Venn diagram showing enrichment of single poly(A) site genes among the differentially expressed genes during mouse cortex development (Fisher's exact test, p=0.00044). (**F**) A Venn diagram showing the overlap between multi-poly(A) site genes that exhibit APA shift and multi-poly(A) site genes that change abundance during mouse cortex development (hypergeometric test, p=0.47).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

PAPERCLIP

Library construction

Sample preparation, immunoprecipitation, SDS-PAGE and RNA extraction were adapted from standard HITS-CLIP (Moore et al., 2014). Mouse monoclonal anti-PABP (Sigma, P6246) was used for immunoprecipitation. The sequencing library was constructed using BrdU-CLIP method (Weyn-Vanhentenryck et al., 2014) with modification to improve sensitivity. The library contains a 14-nt degenerate linker sequence at the 5′ end (6-nt random barcode followed by 8-nt sample multiplexing index) or a 11-nt degenerate linker sequence at the 5′ end (3-nt degenerate sequence, 4-nt sample multiplexing index and 7-nt random barcode). The complete protocol is detailed in Supplemental Protocol. To minimize batch effects, the entire process was performed independently for replicate experiments, sometimes using primers with different indices. Individual PAPERCLIP libraries were multiplexed and sequenced by HiSeq 2000 or MiSeq (Illumina) to obtain 100-nt (HiSeq) or 75-nt (MiSeq) single-end reads.

Read processing and mapping

The processing of raw reads was performed using the CIMS software package previously described (Moore et al., 2014). In brief, the raw reads were filtered based on quality score. Filtered reads with the exact sequence were collapsed into one. Poly(A) sequence at the 3′ end was then trimmed using CutAdapt (Martin, 2011). Only reads that are at least 25-nt in length are retained for mapping to reference genome (hg19 or mm10). Mapping was performed using Novoalign (http://www.novocraft.com) without trimming. A minimum of 25-nt match to the genome sequence was required and only those reads mapped unambiguously to the genome (single hits) were kept for downstream analysis. Reads mapping to the same genomic positions without distinct barcodes were further collapsed into a single tag as previously described (Moore et al., 2014). All reads went through the entire process are referred to as “unique reads” and were used for poly(A) site annotation and other downstream analysis.

Poly(A) site annotation

Unique reads from both replicates were merged. CIMS software package was used to cluster overlapping reads and to determine the read counts (PH) for each cluster. Only clusters that contained more than 2 (for mouse cortex) or 3 (for cultured cells) unique reads were used for downstream analysis. For each filtered cluster, the most abundant 3′ end from all reads in the

cluster was assigned as the poly(A) site. The 3′ most position was selected if there was a tie in abundance for multiple 3′ ends. For cultured cells, poly(A) sites mapped to intergenic regions were excluded from downstream analysis and the remaining poly(A) sites constitute the raw poly(A) site annotation that was used in Figure 1D for comparison with 3′ end sequencing.

To obtain high-confidence poly(A) site annotation in HeLa cells (Figure 2) and HEK293 cells (Figure S1), the raw poly(A) site annotation was further filtered by the following criteria: 1) reproducibility: the poly(A) site had to be present in both replicates within a 20-nt window; 2) abundance: the PH of the poly(A) site had to be at least 10% of gene total. Furthermore, poly(A) sites located > 5kb away from RefSeq 3′ ends or assigned to transcripts that are present in multiple locations in the genome are excluded from the high-confidence set.

Because of the presence of long 3′ UTR extensions (Miura et al., 2013), a modified pipeline was used to annotate poly(A) sites in mouse cortex. The raw poly(A) site annotation was obtained without filtering by genomic location to keep sites from long 3′ UTR extensions. Poly(A) sites within 20kb of RefSeq 3′ ends in the raw annotation were further evaluated for their statistical significance using the tag2peak command in the CIMS software package (Weyn-Vanhentenryck et al., 2014). Statistically significant (with Bonferroni multiple test correction) sites are then further filtered based on reproducibility to obtain the final annotations described in Figure S5.

3′ end sequencing

Library construction

Poly(A) RNA from HeLa cells was prepared using Trizol (Invitrogen) and Dynabeads mRNA Purification Kit for mRNA Purification (Invitrogen). The sequencing library was constructed from 400ng poly(A) RNA using NEBNext mRNA Library Prep Reagent Set for Illumina (New England Biolabs) following manufacturer's instructions with a custom primer for reverse transcription and custom adaptors for ligation (both are adapted from (Jenal et al., 2012)). High-throughput sequencing was performed on MiSeq (Illumina) to obtain 125-nt single-end reads.

Analysis

The processing of raw reads was performed as described in the previous section for PAPERCLIP. Because the library did not contain random barcode, the post-mapping collapse was omitted. Raw poly(A) site annotation was obtained through the same pipeline as for PAPERCLIP described in the previous section.

Comparison to other 3′ end mapping methods

For Figure 1E and 1F, raw reads were downloaded from the NCBI Sequence Read Archive (SRR568012, SRR1033820, SRR299108, SRR090236, SRR453410 and SRR317197) and processed before mapping to hg19 (Novoalign, same settings as for PAPERCLIP). When necessary, poly(A) sequence at the 3′ end was trimmed using CutAdapt. For 3P-Seq and 3′ Seq, two methods that often include untemplated adenines in the sequencing reads, only trimmed reads were used for mapping. For all libraries, only uniquely mapped reads were used for analysis. Downstream adenine-rich reads were defined as in the literature (Derti et al., 2012; Shepard et al., 2011): reads that have 6 consecutive adenines or at least 7 adenines total in the 10 nucleotides immediately downstream of the last aligned position in the human genome. Reference BED files were downloaded from the UCSC genome browser for annotating the genomic localization of mapped reads.

RNA-seq and analysis

Library construction

Poly(A) RNA from HeLa cells and adult/embryo mouse cortex was prepared using Trizol (Invitrogen) and Dynabeads mRNA Purification Kit for mRNA Purification (Invitrogen). Stranded RNA-seq libraries were constructed from 50ng poly(A) RNA using ScriptSeq V2 kit (Epicentre) following manufacturer's instructions. High-throughput sequencing was performed on MiSeq (Illumina) to obtain 35-nt or 37-nt paired-end reads.

Analysis

For Figure S2, reads were mapped to the annotated human transcriptome (GRCh37) using tophat2 (version 2.0.11). Duplicated reads were then removed using Picard tools (1.107) and the resulting bam file was converted to a bed file. The bed file in turn was used to create the bedgraph. For Figure S5D, reads were aligned to the mouse genome (mm10) using bowtie2 (version 2.1.0) and converted the resulting sample to a bam file containing only the properly mapped paired-reads. The bam file was then converted to a paired-end bed file using bed tools (version 2.20.1), which was in turn used to create a bed graph file (tag2profile, CIMS package). For Figure S6E-F, reads were mapped to the mm10 genome assembly using STAR aligner using default parameters and providing splicing junctions (Ensembl release 80). HTSeq package (Anders et al., 2015) was used to count the number of overlapping reads per gene definition. Replicate raw gene read counts were used as input to EdgeR for differential gene expression analysis.

PAPERCLIP analysis of APA shift

Poly(A) site usage for each library was determined by the count of reads aligning up to 100-nt upstream of a PAPERCLIP identified poly(A) site. Genes with poly(A) sites closer than 100-nt apart were not considered in the APA shift analysis. Also, we restricted the APA shift analysis to genes with two distinct PAPERCLIP defined poly(A) sites. Additionally, for the HeLa siRNA comparisons, at least one of the poly(A) site was required to be present in the control experiment. EdgeR package (Robinson et al., 2010) was used to statistically test significant APA shifts between two experimental conditions, while accounting for biological and technical variability between experimental replicates. Each replicate dataset was first normalized to account for library size and compositional bias using the TMM methodology (Robinson and Oshlack, 2010). Next, we model the poly(A) site PAPERCLIP read count data as a negative binomial distribution and fitted a Generalized Linear Model (GLM) with explanatory variables for batch, experimental condition (e.g. siRNA treatment, mouse developmental stage), poly(A) location (i.e distal or proximal) and an interaction factor for experimental condition × poly(A) location. For each gene, GLM likelihood ratio test was conducted to test if the interaction coefficient between experimental condition and poly(A) location was non-zero (i.e poly(A) site usage is dependent on experimental condition). GLM likelihood ratio test derived p-values were adjusted for multiple hypotheses testing using the qvalue package (Storey and Tibshirani, 2003). In addition to FDR < 0.05 cutoff, we require the average normalized read count ratio of poly(A) locations (proximal vs. distal) to be greater than a twofold change between experimental conditions to identify genes with robust APA shifts.

Motif analysis

Nucleotide word enrichment for all possible 4-mers and 6-mers was computed from input FASTA files as previously described (Licatalosi et al., 2012). To construct the control sequences, the query regions were shifted 500bp in the 5′ direction (for the proximal poly(A) site) or in the 3′ direction (for the distal poly(A) site). *De novo* motif discovery was performed using DREME (Bailey, 2011).

Ago HITS-CLIP and target analysis

Highly reproducible (biological complexity ≥ 6 out of 15) Ago HITS-CLIP clusters from the combination of both previously described (Chi et al., 2009) and newly generated data (GSE73058) (Moore et al., 2015) in mouse cortex were used for analysis. Targets were predicted using the miRanda algorithm (Enright et al., 2003) for the top 100 most abundant

miRNAs in the Ago HITS-CLIP dataset. miRanda was run with default settings with minimum target score of 140, which requires both high level of seed pairing and thermodynamic stability of the miRNA:mRNA duplex.

Novel miR-128 target identification

GEO dataset GSE48813 was downloaded from NCBI and used to calculate the fold-change between miR-128 knockout and control groups. The original p-value included in the dataset was used to determine the statistical significance for the fold-change. A gene needs to satisfy the following criteria to be considered as a novel miR-128 target: 1) the presence of miR-128 binding site in the extended 3′ UTR region but not in the upstream region; 2) statistically significant ($p \leq 0.05$) increased ribosome-association (fold-change >1) in miR-128 deficient D1-neurons for at least one probe set in GSE48813. Wilcoxon rank sum test was used to evaluate statistical significance between groups in Figure 5D.

Comparison between APA shift and DERs

Raw RNA-seq reads from 6 fetuses and 6 adults were downloaded from NCBI (Accession code: PRJNA245228). Reads were mapped to hg19 using STAR (Dobin et al., 2013) and the resulting bam file was converted to a bed file. Reads mapped to the upstream exons, the proximal 3′ UTR and distal 3′ UTR were counted using BEDtools. For quantifying the relative usage of *Atp2b2* proximal and distal 3′ UTRs, the proximal 3′ UTR-to-upstream exons ratio, which is similar between individuals (except for adult SRR1554535.2 and SRR1554536.3), serves as an internal control. For Figure 6E, data from the two aforementioned adults were excluded to give a conservative estimate. The distal 3′ UTR-to-proximal 3′ UTR ratio is actually higher if all adults were included.

Cell-type expression analysis in mouse brain

For 444 genes described in Figure 6A, FPKM data from a previously published RNA-seq study (Zhang et al., 2014) are used to determine their possible enrichment in specific cell-types. Fold-enrichment is calculated as FPKM of one cell type divided by the average FPKM of all other cell types. The average FPKM from the three oligodendrocyte-lineage cells is used to represent the oligodendrocyte lineage FPKM. Genes with fold-enrichment >=5 are considered enriched in specific cell types.

Northern blots

Total RNA from HeLa cells was prepared by Trizol (Invitrogen) extraction and column purification using High Pure RNA Isolation Kit (Roche). 10µg total RNA per lane was separated with 0.8% SeaKem Gold agarose (Lonza) along with Millennium RNA Markers (Ambion). The gel was treated with 0.05M NaOH; 1.5M NaCl for 20 min., then with 0.5M Tris, pH7.5; 1.5M NaCl for 10 min., before being equilibrated for blotting in 20xSSC for 20 min. The gel was blotted onto Hybond-N filter (Amersham). Probes were prepared using Prime-It II Random Primer Labeling Kit (Agilent) with [α-^{32}P] dCTP. All primer sequences are listed in Table S1. For hybridization, labeled probes were used at 1 x 10^6cpm/ml of hybridization solution (1% BSA Fraction V, 7% SDS, 0.5M NaH_2PO_4, pH 7, 1mM EDTA, pH 8). The filter was hybridized overnight at 68ºC and washed with 2xSSC; 0.05%SDS at 60ºC 2x 10min., then with 0.1xSSC; 0.1%SDS at 60ºC 2x 30min.

Cell culture

HeLa and LN229 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% FBS. For Figure 3 and S3, 900,000 HeLa or LN229 cells were plated in 10-cm dishes the day before siRNA transfection. Transfection was performed using DharmaFECT1 (Dharmacon) following standard protocol provided by the manufacturer. The following pre-designed siRNAs were used: si-CFIm68 (s21773 and s223186, Ambion), si-CstF64/64tau (s3684, s3686, s23471 and s23472, Ambion) and si-Ctrl (On-TARGETplus non-targeting pool, Dharmacon). Individual siRNAs were used at final concentration of 5nM and the total concentration of siRNA pool is 10nM (si-CFIm68) or 20nM (si-Ctrl and si-CstF64/64tau). 72 hours after transfection, the transfected ells were UV-crosslinked for PAPERCLIP experiments or harvested for protein and RNA. For Figure 4E, 1.2 million HeLa cells were plated in 10-cm dishes the day before plasmid transfection. Transfection was performed with 3µg plasmid and 18µl X-tremeGENE 9 (Roche) per dish. For Figure 4H, S4C, S4E and S4H, 150,000 HeLa cells were plated in 6-well plates the day before siRNA transfection (day 1). The next day (day 2), siRNA transfection was performed using DharmaFECT1 (Dharmacon) following standard protocol provided by the manufacturer. The following day (day 3), the transfected cells were split 1:2 and re-plated in 6-well plates. On day 4, plasmid transfection was performed with 0.5µg plasmid and 3µl X-tremeGENE 9 (Roche) per well. The cells were harvested for RNA on day 5.

3′ RACE

Poly(A) RNA from HeLa cells was prepared using Trizol (Invitrogen) and Dynabeads mRNA Purification Kit for mRNA Purification (Invitrogen). 40ng poly(A) RNA was used for cDNA

synthesis using an anchored primer and SuperScript III (Invitrogen) following manufacturer's instructions. Gene-specific forward primers and a common reverse primer were used for the first-round and nested PCR amplification. All primer sequences are listed in Table S1. The nested PCR products were cloned using Zero Blunt TOPO PCR Cloning Kit (Invitrogen) following manufacturer's instructions and sequenced by Sanger sequencing.

SDS-PAGE and western blots

20μg total protein per lane was separated on 10% Novex NuPAGE Bis-Tris gels (Invitrogen) and transferred to nitrocellulose membrane following standard procedures. The following antibodies are used for western blotting: rabbit polyclonal anti-CFIm68 (A301-356A, Bethyl Laboratories; 1:1000), rabbit polyclonal anti-CstF64 (A301-093A, Bethyl Laboratories; 1:1000), rabbit polyclonal anti-CstF64tau (A301-487A, Bethyl Laboratories; 1:1000), and mouse monoclonal anti-GAPDH (ab8245, Abcam; 1:2000).

qRT-PCR

For Figure 3E and 3F, RNA from siRNA-transfected HeLa cells was prepared using High Pure RNA Isolation Kit (Roche) with on-column DNase I digestion. 5μg total RNA was used for cDNA synthesis using an anchored primer and SuperScript III (Invitrogen) following manufacturer's instructions. For Figure 6B, 6C, S3E and S3F, RNA from adult/embryo mouse cortex or LN229 cells was prepared using Trizol (Invitrogen) and Dynabeads mRNA Purification Kit for mRNA Purification (Invitrogen). 40ng poly(A) RNA was used for cDNA synthesis using the same anchored primer and SuperScript III (Invitrogen) following manufacturer's instructions. For Figure 4H, S4C, S4E and S4H, RNA from transfected HeLa cells was prepared using Trizol (Invitrogen). 1~1.5μg total RNA was first digested with DNase I (Invitrogen) and then used for cDNA synthesis using an anchored primer and SuperScript III (Invitrogen) following manufacturer's instructions.

qRT-PCR was performed using FastStart SYBR Green Master mix (Roche) in triplicates. All primer sequences are listed in Table S1. The cycling parameters are: 95ºC for 10 min. followed by 40 cycles of 95ºC for 15 sec., 58ºC for 30 sec., 72ºC for 20 sec. Quantification was calculated using the ΔΔCt method with the following endogenous controls: ACTB (HeLa) and Rpl13a (mouse) or using the standard curve method with serially diluted pEN2 plasmid DNA as the standards. For Figure 3E and 3F, $\log_2$(distal/proximal ratio) in si-CFIm68 or si-CstF64/64tau relative to $\log_2$(distal/proximal ratio) in si-Ctrl was plotted for each gene. For Figure 6B and 6C,

$\log_2$(distal/proximal ratio) in adult relative to $\log_2$(distal/proximal ratio) in embryo was plotted for each gene.

Animal

Adult (8-14wk) and pregnant (E15.5) C57BL/6J mice were obtained from the Jackson Laboratory. Dissection of adult and embryonic brain cortex was performed following standard procedures. All procedures were conducted according to the Institutional Animal Care and Use Committee (IACUC) guidelines at the Rockefeller University.

Luciferase reporter construction and assay

For all candidate genes, neuronal FPKM data from a previously published RNA-seq study (Zhang et al., 2014) are used to determine their expression in neurons. All 8 genes in Figure 5A have FPKM > 0.1, the statistically significant level described in the original study. To construct luciferase reporter, putative miR-128 binding sites with ~200nt flanking sequence were amplified from mouse genomic DNA and inserted into psiCheck2 vector (Promega) between the PmeI and NotI sites. The genomic coordinates of miR-128 binding sites are listed in Table S5. Q5 Site-Directed Mutagenesis Kit (New England Biolabs) was used to create mutant miR-128 binding sites. All primer sequences are listed in Table S1. The PmeI-NotI fragments containing the mutant binding sites were subcloned back to psiCheck2 following site-directed mutagenesis to create mutant luciferase reporters. All insert sequences were verified by Sanger sequencing.

For luciferase assay, 60,000 HEK293 cells were plated in 24-well plates the day before transfection. For each well, 100ng psiCheck2 reporter DNA and 12.5 pmole microRNA mimic (final concentration: 25nM) were co-transfected using DharmaFECT Duo (Dharmacon) following standard protocol provided by the manufacturer. The following microRNA mimics from Dharmacon were used: mmu-miR-128-3p (C-310398-07), mmu-miR-137-3p (C-310413-05) and miRIDIAN microRNA mimic negative control (CN-001000-01). 24 hours after transfection, cells were lysed and assayed for firefly and *Renilla* luciferase activity in triplicate using the Dual-Luciferase Reporter Assay System (Promega) on a luminometer at the Rockefeller University High-Throughput and Spectroscopy Resource Center. Transfection efficiency is controlled by normalizing *Renilla* luciferase activity to firefly luciferase activity. Mutant reporter activities were further normalized to wildtype reporter activities to obtain relative luciferase activities.

pEN2 and related constructs

pEN2 was created in the following steps: 1) Remove the SV40 poly(A) site from pAcGFP-C1 (Clontech) by ligation of filled-in product from MluI-HpaI digestion. 2) TOPO cloning of the entire human EN2 3′UTR (PCR-amplified from human genomic DNA). 3) Subclone the EN2 3′UTR into modified pAcGFP-C1 between BglII and EcoRI sites. Q5 Site-Directed Mutagenesis Kit (New England Biolabs) was used to create point mutations and all mutants were subcloned back to the original backbone following site-directed mutagenesis. For pEN2-SRSF9, pEN2-JUNB and the three NOTCH constructs described in Fig. S4E, oligonucleotides were annealed and directly ligated into a modified pEN2 construct in which SpeI and XhoI sites were introduced to facilitate cloning. pAcGFP-JUNB was created by inserting the entire JUNB 3′ UTR plus ~100bp sequence 3′ to the JUNB poly(A) site into pAcGFP-C1 between XhoI and EcoRI sites. All insert sequences were verified by Sanger sequencing. All primer sequences are listed in Table S1.

SUPPLEMENTAL REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486.

Derti, A., Garrett-Engele, P., MacIsaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. Genome Res *22*, 1173–1183.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Oude Vrielink, J.A.F., Bos, A.J., Drost, J., Rooijers, K., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell *149*, 538–553.

Licatalosi, D.D., Yano, M., Fak, J.J., Mele, A., Grabinski, S.E., Zhang, C., and Darnell, R.B. (2012). Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. Genes Dev *26*, 1626–1642.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal.

Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O., and Lai, E.C. (2013). Widespread and extensive lengthening of 3' UTRs in the mammalian brain. Genome Res *23*, 812–825.

Moore, M.J., Scheel, T.K.H., Luna, J.M., Park, C.Y., Fak, J.J., Nishiuchi, E., Rice, C.M., and Darnell, R.B. (2015). miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. Nat Comms *6*, 8864.

Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nature Protocols *9*, 263–293.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and

dynamic landscape of RNA polyadenylation revealed by PAS-Seq. Rna *17*, 761–772.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci USA *100*, 9440–9445.

Tan, C.L., Plotkin, J.L., Venø, M.T., Schimmelmann, von, M., Feinberg, P., Mann, S., Handler, A., Kjems, J., Surmeier, D.J., O'Carroll, D., et al. (2013). MicroRNA-128 governs neuronal excitability and motor behavior in mice. Science *342*, 1254–1258.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Reports *6*, 1139–1152.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. J Neurosci *34*, 11929–11947.

SUPPLEMENTAL PROTOCOL

PAPERCLIP procotol, total of 12 pages (see next page)

PAPERCLIP [*P*oly(*A*) binding *P*rotein-mediated mRNA 3´*E*nd *R*etrieval by *C*ross*L*inking *I*mmuno*P*recipitation]

Day 1

I. General method for UV cross-linking of tissue/cell lines

a. UV cross-linking

For mouse tissue:

Harvest tissue and let tissue sit in ice cold 1x PBS (or 1x HBSS) until harvest is complete.

[We used 150mg mouse cortex, which is a good starting point for brain tissue. The optimal amount of tissue needs to be tested.]

Add ~10 tissue volumes of 1x PBS (or 1x HBSS) and triturate tissue with 18-gauge needle and syringe a few times.

[Because UV light can penetrate a few cell layers, stringent trituration to the single-cell is not necessary.]

Spread the tissue suspension in 10-cm petri dishes and put the dishes on ice in a tray with the lid off. Put the tray in Stratalinker 2400 (Stratagene) and irradiate three times for 400mJ/cm^2. Rotate the tray 90° between each irradiation.

[The length of crosslinking may need to optimized for other types of tissue or a different UV-crosslinker is used. For a preliminary experiment, try 100, 200 and 400mJ/cm^2, and then use the shortest condition that gives >70% of the maximum signal.]

For cell culture:

Grow cells in a 10-cm or 15-cm dishes, remove growth media, wash once with ice-cold 1x PBS, add a thin layer of ice-cold 1x PBS. Put the dishes on ice in a tray with the lid off. Put the tray in Stratalinker and irradiate three times for 400mJ/cm^2. Rotate the tray 90° between each irradiation.

[We crosslink HeLa and HEK293 cells in 4mL 1xPBS per 10-cm dish. The length of crosslinking may need to optimized for other types of cells or a different UV-crosslinker is used. For a preliminary experiment, try 100, 200 and 400mJ/cm^2, and then use the shortest condition that gives >70% of the maximum signal.]

b. Post-crosslinking processing

Scrape the cell monolayer in 1x PBS, transfer the cell/tissue suspension to 1.5mL or 2mL eppendorf tubes, pellet at 6000 rpm for 3 min at 4°C, remove supernatant and freeze pellets at -80°C until use.

[For convenience, each eppendorf tube will be used in one reaction for the following steps. We put ~6 million HeLa cells or ~150mg mouse cortex in one eppendorf tube.]

To prepare **1x HBSS**:

50ml 10x Hank's Balanced salt solution, Ca-Mg-free (Invitrogen, #14185-012)
5ml 1M HEPES, pH 7.3
445ml ddH_2O

Day 2

II. Immunoprecipitation

a. Solutions

Antibody Binding Buffer
1x PBS (tissue culture grade; no Mg++, no Ca++)
0.02% Tween-20

1x TS Buffer
1x PBS (tissue culture grade; no Mg++, no Ca++)
0.1% SDS
1.0% Triton X-100

2x TS Buffer
2x PBS (tissue culture grade; no Mg++, no Ca++)
0.1% SDS
1.0% Triton X-100

1x PNK Wash Buffer
50mM Tris-Cl pH 7.4
10mM $MgCl_2$
0.5% NP-40

1x PNK+EGTA Wash Buffer
50mM Tris-Cl pH 7.4
0.5% NP-40
20mM EGTA

b. Bead preparation

For each eppendorf tube of crosslinked tissue (~150mg mouse cortex or 6 million HeLa cells), use 200~300µl of Dynabeads Protein G (Invitrogen, 10004D).

[Prepare enough beads for all reactions. The minimum number of reactions is two (one high RNase and one low RNase).]

Wash beads 3x with Antibody Binding Buffer.

Re-suspend beads in 200~300µl Antibody Binding Buffer and add 30~40µg of 10E10 anti-PABP (Sigma, P6246).

Rotate beads at room temperature for at least 30 minutes.

Wash beads three times with 1x TS Buffer; if you are not yet ready to add crosslinked lysate, leave beads in last wash step on ice.

c. Prepare crosslinked lysate

Resuspend each tube of crosslinked tissue using 500µl~1mL 1x TS Buffer (>5x tissue volume); sit on ice for 5-10 min to lyse.

Add 30~40 µl of RQ1 DNase (Promega, M6101) to each tube; incubate at 37° for 5 min, 1000 rpm.

[We perform all incubation/shaking steps in Eppendorf Thermomixer R. A thorough and even shaking is critical.]

Make a dilution of RNase A (USB, 70194Y) at 1:100 in 1x TS Buffer (High-RNase).

Make a dilution of RNase A (USB, 70194Y) at 1:100,000 in 1x TS Buffer (Low-RNase).

[Each experiment should be done in duplicate with two RNAse concentrations – the dilution depends also on the batch of RNAse, so in the first experiment, several dilutions should be tested. The High-RNase group is used as a control to confirm that the size of the radioactive band on SDS-PAGE gel changes in response to different RNAse concentrations (which confirms that the band corresponds to a protein-RNA complex).]

Add 10~20µl of 1:100 diluted RNase A per 1mL lysate to the High-RNase tube.

Add 10µl of 1:100,000 diluted RNase A per 1mL lysate to the Low-RNase tube(s).

Incubate at 37°C for 5 minutes, 1000 rpm.

[Optimal amount of RNase A needs to be experimentally determined.]

Spin lysates in pre-chilled microcentrifuge at 14,000 rpm (Max, ~20,800g) for 10 minutes at 4°C.

Remove the supernatant from the precipitates.

d. Immunoprecipitation

Divide washed beads; add supernatant to one prepared tube of beads.

Rotate beads/lysate mix for 2 hours at 4°C.

e. Post-IP washes

Wash beads in the following order (all buffers are ice-cold and kept on ice during washes):
-Two times with 1x TS Buffer
-Two times with 2x TS Buffer
-Two times with 1x PNK Wash Buffer

III. CIP Treatment, (On-Bead)

CIP mix:

6 μl 10x Fast AP Buffer
2.3 μl Fast AP (Fermentas, EF0654)
0.6 μl 10% Tween-20
51.1 μl water
60 μl total

[The reaction volume is based on 300µl protein-G beads and can be scaled down if fewer amounts of beads are used.]

Add 60 μl of PNK mix to each tube and incubate in Thermomixer R (Eppendorf) at 37° for 20 minutes (1200 rpm every 1.5 minutes for 15 seconds).

Wash:
-Once with 1x PNK Wash Buffer
-Once with 1x PNK+EGTA Wash Buffer
-Two times with 1x PNK Wash Buffer

IV. 5′ labeling, (On-Bead)

PNK mix:

6 μl	10x PNK Buffer (NEB)
3 μl	T4 PNK (NEB, M0201L)
1.5 μl	^{32}P-γ-ATP
49.5 μl	water
60 μl	total

[The reaction volume is based on 300µl protein-G beads and can be scaled down if fewer amounts of beads are used.]

Add 60 μl of PNK mix to each tube and incubate in Thermomixer R (Eppendorf) at 37° for 20 minutes (1200 rpm every 1.5 minutes for 15 seconds).

Wash:
-Three times with 1x PNK Wash Buffer

Leave beads in the last wash and store at 4°C overnight.

[It is possible to complete all Day 2 and Day 3 procedures in a very long day. In this case, one can proceed to the next part without storing the reactions.]

Day 3

V. SDS-PAGE & nitrocellulose transfer

a. Elution

Elution mix:

27 μl 1x PNK Buffer
3 μl 1M DTT
30 μl NuPAGE 4x LDS Sample Buffer (Invitrogen, NP0007)
60 μl total

Remove the last 1x PNK Wash Buffer; add 60 μl elution mix to each tube.

Elute Protein/RNA complexes from beads by incubating at 70°C for 10 minutes (1200rpm).

Load 1 tube per 2 wells (30 μl/well) of a 20 well Novex NuPAGE 10% Bis-Tris Midi gel.

Run the gel at 200V in the cold room. (Run time: ~1.5hrs with 1x MOPS buffer)

[Some pre-stained molecular weight markers may run differently on Novex NuPAGE gels. We use rainbow marker (GE Healthcare, RPN800E), which runs at the expected molecular weights.]

After gel run, transfer gel to BA-85 nitrocellulose membrane using the Novex wet transfer apparatus.

[This pure nitrocellulose is a little fragile, but it works better for the RNA/protein extraction step.]

Transfer at 30V in NuPAGE Transfer Buffer with 10% methanol. (Transfer time: 75~90 minutes)

After transfer, rinse the nitrocellulose membrane in 1x PBS, and gently blot on Kimwipes; wrap membrane in plastic wrap and expose to film in a cassette at -80°C.

[Use a luminescent sticker, so that you can later align the membrane back to the autoradiogram. A band a little bit above the 76kDa marker should be readily observed for the High-RNase group. (The molecular weight of PABP is ~75kDa)]

[The exposure time is usually 1~4 hours when using fresh ^{32}P-γ-ATP. Reactions that require exposure times longer than 4 hours often result in low complexity libraries.]

Day 4

VI. RNA Isolation and Purification

1x PK Buffer:

100 mM	Tris-Cl pH 7.5
50 mM	NaCl
10 mM	EDTA

1x PK Buffer/7M urea (this buffer must be fresh):

100 mM	Tris-Cl pH 7.5
50 mM	NaCl
10 mM	EDTA
7 M	Urea

Cut nitrocellulose membrane (bottom margin: slightly above the top of the High-RNase band; top margin: ~225kDa marker) using a clean scalpel blade, and cut into small pieces (the smaller the better). Put the nitrocellulose pieces into 1 eppendorf tube for each group.

Make a 4mg/ml proteinase K (Roche, 03115828001) solution in 1x PK Buffer; pre-incubate this stock at 37°C for 10-20 minutes to kill any RNases.

Add 200 μl of proteinase K solution to each tube of isolated nitrocellulose pieces; incubate 20 min at 37°C at 1200 rpm.

Add 200 μl 1x PK/7M urea solution; incubate another 20 min at 37°C at 1200 rpm.

Add 400 µl RNA phenol (Sigma, P4682-100ML) and 130 µl of $CHCl_3$ (Sigma, 25668-100ML) to solution; vortex at high for 20 seconds; incubate 20 min at 37°C at 1200 rpm.

[RNA phenol can also be prepared by equilibrating pure phenol with 0.15 M NaOAc pH 5.2; $CHCl_3$ is chloroform 49:1 with isoamyl alcohol.]

Spin tubes at full speed in microcentrifuge; transfer aqueous phase to new eppendorf tubes.

Add the following to each tube:

50 μl	3M NaOAc, pH 5.2
1 μl	Ultra pure glycogen (Invitrogen, 10814-010)
1 ml	1:1 mix of ethanol and isopropanol

Precipitate overnight at -20°C.

Day 5

VII. cDNA synthesis and purification

a. Bead Preparation: Blocking with Denhardt's Solution

Ab Binding Buffer:
1X PBS, pH 7.4
0.02% Tween-20

50µl Protein-G Dynabeads per sample (25µl per cDNA purification step), include -RT and/or -Template

Wash 3 times with Ab binding buffer

Add 225µl Ab binding buffer, 25µl 50X Denhardt's Solution (Sigma, D2532 or Invitrogen, 750018); total volume is 5X original bead volume

Rotate at RT for at least 45 minutes ~ 1 hour

b. Reverse Transcription

Spin down the RNA at max speed (14,000 rpm, ~20,800g) for 20 minutes at 4°C. Wash 2 times with 75% ethanol, and dry the pellet.

Add 8µl nuclease-free water to RNA pellet (tap to resuspend, quick spin down). Denature at 65°C for 5 minutes (in microfuge tube), place tube on ice (to avoid loss of RNA, do not over-dry pellet and do not pipette until after denaturing step)

Transfer to PCR tube (on ice)

Mix I:
4µl 5X RT Buffer
1µl dATP
1µl dCTP } 8.2mM (Invitrogen, 10297-018)
1µl dGTP
1µl Br-dUTP (8.2mM; Sigma, B0631)
8µl total

Add 8µl of Mix I, 1µl of 25µM indexed RT primer to each tube.

Denature 3 minutes at 75°C, ramp down to 48°C and hold.

Mix II:
1µl 0.1M DTT
1µl RNasin Plus (Promega, N2611)
1µl SuperScript III (Invitrogen)
3µl total

Add 3µl of Mix II (pre-warm to 48°C in PCR block before adding).

Reverse transcription: 45 minutes at 48°C, 15 minutes at 55°C, 5 minutes at 85°C, 4°C hold.

c. Bead Preparation: Antibody binding

1x IP Buffer:
0.3X SSPE
1mM EDTA
0.05% Tween-20

Wash blocked beads 3 times with Ab binding buffer.

Add 5µl 50X Denhardt's Solution, 5~7.5µl (5~7.5µg) anti-BrdU antibody (Millipore, MAB3222), and 37.5~40µl Ab binding buffer to bring the total reaction volume to 50µl.

Rotate at RT for at least 45 minutes.

Wash 3 times with 1x IP Buffer.

d. Post-RT clean-up:

Add 1µl (at 2U/µl) RNase H (Invitrogen 18021-071 or NEB M0297L) to each RT tubes.

Incubate for 20 minutes at 37°C, hold at 4°C.

Add 10µl nuclease-free water (to bring volume above 25µl needed for G-25 column).

Spin the RNase H-digested RT products through Illustra Microspin G-25 column (GE Healthcare, 27-5325-01) to remove free BrdUTP (discard G-25 column as solid radioactive waste).

e. cDNA Purification: Immunoprecipitation I

2x IP Buffer:
0.6X SSPE
2mM EDTA
0.1% Tween-20

Nelson Low Salt Buffer:
15mM Tris pH 7.5
5mM EDTA

Nelson Stringent Buffer:
15mM Tris-HCl pH7.5
5mM EDTA
2.5mM EGTA
1% Triton X-100
1% Sodium deoxycholate
0.1% SDS
120mM NaCl
25mM KCl

Measure volume, add water up to 40µl and add 10µl 50X Denhardt's Solution and 50µl 2x IP Buffer for a total volume of 100µl (Denhardt's and 2X IP Buffer can be added to the G-25 column collection tube prior to spinning samples through, volume can then be adjusted up to 100µl).

Denature 5 minutes at 70°C, equilibrate to room temperature.

Add to prepared tube of beads (25µl original slurry volume, store remaining beads for second purification at 4°C overnight), rotate at RT for 45 minutes.

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 1 time with 1x IP Buffer (5x Denhardt's)
2 times with Nelson Low Salt Buffer (1x Denhardt's)
2 times with Nelson Stringent Buffer (1x Denhardt's)
2 times with 1x IP Buffer

f. cDNA Purification: Heat Elution

Elution Buffer:
50µl 2x IP Buffer
40µl Water
90µl

Add 90µl elution buffer to each tube of beads.

Incubate at 98°C for 1 minute in thermomixer (1200 rpm) to elute cDNA from beads and place on magnet

Collect eluate and add 10µl 50X Denhardt's to each tube for a total of 100µl

Store overnight at 4°C

Day 6

VIII. cDNA purification and library construction

a. cDNA Purification: Immunoprecipitation II

CircLigase Wash Buffer:
33mM Tris-Acetate
66mM KCl
(pH 7.8)

Denature 5 minutes at 70°C, equilibrate to room temperature.

Add to prepared tube of beads, rotate at RT for 45 minutes.

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 1 time with 1x IP Buffer (5x Denhardt's)
2 times with Nelson Low Salt Buffer (1x Denhardt's)
2 times with Nelson Stringent Buffer (1x Denhardt's)
2 times with CircLigase Wash Buffer

b. cDNA Circularization with CircligaseII

Phusion Wash Buffer:
50mM Tris
(pH 8.0)

Circligase Reaction Mix:
2µl CircLigase 10X Reaction Buffer
4µl Betaine (5M)
1µl $MnCl_2$ (50mM)
1µl CircLigase ssDNA Ligase II (100U) (Epicentre, CL9021K)
12µl Water
20µl total

Incubate 1 hour at 60°C in thermomixer (interval: shake at 1300rpm every 30" for 15").

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 2 times with Nelson Low Salt Buffer
2 times with Nelson Stringent Buffer
2 times with Phusion wash buffer

c. PCR: Phusion Polymerase, SYBR Green

Mix I:
10µl 5X Phusion HF Buffer
1µl 10mM dNTPs
37µl Water
48µl total

Mix II:
0.5µl DP5-PE (20µM)
0.5µl DP3-PAT (20µM)
0.5µl Phusion DNA Polymerase (NEB, M0530)
1.5µl total

Add 48µl Mix I to beads.

Incubate at 98°C for 1 minute in thermomixer (1200 rpm) to elute cDNA from beads and place on magnet.

Collect eluate and place in PCR tube with optically clear cap.

Add 1.5µl Mix II, 0.5µl 50X SYBR Green I (dilute 10,000X stock to 50X in Phusion Wash Buffer) to mix and place in real-time PCR machine (Bio-Rad, CFX96).

(PCR cycle parameters)

Initial denaturation:	98°C	30”
Cycle:	98°C	10”
	60°C	15”
	72°C	20”

Remove reaction tube when RFU signal reaches ~250-500 (usually results in 2.5-5nM).

e. Post-PCR processing: library purification and quantitation

Purify PCR product using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer’s instructions.

Quantitate using TapeStation (Agilent), pool samples according to TapeStation results.

MiSeq:
Dilute to 2nM, sequence on MiSeq at 5pM final concentration using standard Read 1 primer already on cartridge.

HiSeq:
Dilute to 2-10nM Submit for HiSeq sequencing using standard Read 1 sequencing primer listed below.

IX. Primers

RT Primers:

RT-1 (CGAT):

/5Phos/DDDATCGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CAA
GCAGAAGACGGCATACGAGATTTTTTTTTTTTTTTTTTTVN

RT-2 (TAGC):

/5Phos/DDDGCTANNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CAA
GCAGAAGACGGCATACGAGATTTTTTTTTTTTTTTTTTTVN

RT-3 (CTAG):

/5Phos/DDDGACTNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CAA
GCAGAAGACGGCATACGAGATTTTTTTTTTTTTTTTTTTVN

RT-4 (GATC):

/5Phos/DDDAGTCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CAA
GCAGAAGACGGCATACGAGATTTTTTTTTTTTTTTTTTTVN

PCR Primers:

DP5-PE: (Allowing standard Read1 sequencing primer)

5′–AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

DP3-PAT:

5′–CAAGCAGAAGACGGCATA

Sequencing Primers:

Illumina Standard Read1 Sequencing Primer:

5′–ACACTCTTTCCCTACACGACGCTCTTCCGATCT